DOCUMENT RESUME

ED 473 109 HE 035 648

AUTHOR	Liu, Richard
TITLE	A Methodological Critique of Tinto's Student Retention Theory. AIR 2002 Forum Paper.
PUB DATE	2002-06-00
NOTE	25p.; Paper presented at the Annual Forum for the Association for Institutional Research (42nd, Toronto, Ontario, Canada, June 2-5, 2002).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Persistence; Causal Models; *College Students; Dropout Research; *Dropouts; Higher Education; Path Analysis; *Student Attrition
IDENTIFIERS	Tinto (V); *Tinto Theory

ABSTRACT

College student retention has become the area of research that has done the most to integrate various administrative factors and academic disciplinary concerns in the research agenda of higher education. V. Tinto's model has long been cited as the major theory in explaining dropout behavior. As his theory is so intertwined with path analysis, empirical researchers have needed to examine some major concerns, which have previously been largely unintended. The main purpose of this essay is to explore Tinto's model in the context of causal modeling and to discuss its methodological difficulties and ramifications and ensuing issues. Empirical studies using Tinto's theory are cited for illustration. (Contains 6 figures and 49 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 473 109

A Methodological Critique of Tinto's Student Retention Theory

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D.Vura

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

by

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Richard Liu
University of Minnesota Duluth

Correspondence should be sent to Dr. Richard Liu, Institutional Research, 420 Darland Admin Bldg , 10 University Drive. Duluth Mn 55812

Phone: 218-726-6832
e-mail address:rliu@d.umn.edu

BEST COPY AVAILABLE

HE 035 648

A Methodological Critique of Tinto's Student Retention Theory

Abstract

College student retention has become the area of research that has done the most to integrate various administrative factors and academic disciplinary concerns in the research agenda of higher education. Tinto's model has long been cited as the major theory in explaining dropout behavior. As his theory is so intertwined with path analysis, empirical researchers have needed to examine some major concerns, which have previously been largely untended. The main purpose of this paper is to explore (1) Tinto's model in the context of causal modeling, (2) its methodological difficulties and ramifications, and (3) ensuing issues. Empirical studies using Tinto's theory were cited for illustration.

The easy use of the multiple regression procedure and its availability in software greatly encouraged social sciences in their research efforts and facilitated the positivistic approach, as opposed to the humanistic approach, in the late sixties. More important in the positivistic approach was the development of Path Analysis in the seventies. Steinschombe's construction of social theory provided the lineage of the path technique, which led to the construction of sociological theory (1968). Causal modeling then became an extremely popular tool of sociologist to use in exploring a sociological research agenda. The combination of (1) the statistical technique of Path Analysis and (2) theory building provided an impetus for philosophical investigation.

Casual analysis was not a new technique, yet when it was mingled with statistical analysis its theoretical foundation was solidified. What social scientists were arguing was not whether there is a final or permanent cause and effect relationship between events, but rather whether within a set of time references a causal relationship between variable x and y can be established. The statement that there are regular connections between certain events or qualities is an empirical one, for such connections are observable. The assertions that (1) events are connected as a matter of fact, and (2) one event necessitates the occurrence of the other event are discarded (Zilsel, 1968). Causal analysis' popularity among social scientists hinted at its universal appeal among academics in diverse fields of interest. Vienna Circle was not mentioned in the movement yet its influence was all too visible. That the theoretical postulate is amenable to objective verification placed Path Analysis prominently in social science research, especially in sociology. A cross-discipline approach to problems was deemed necessary and appropriate. Path Analysis, first used by Wright (1921) in biology now became a common ground shared by the fields of sociology, psychology, political science and econometrics.

On a separate front, System Analysis espoused by Parsons and Merton elevated sociology to a new status. Merton's analysis of Social Structure and Anomie was a combination of an application of Durkheim's sociology as well as system analysis (Friedrichs, 1972). Anomie, along with the antonym of integration, had become prevalent terms in the sociology of deviance as well as the sociology of knowledge. Durkheim's influence upon American sociology reached its peak in the early seventies and continued to spill over into the field of higher education, first through Spady's work and later through Tinto's work. Both Spady and Tinto were sociologists by training and thus the influence of Durkheim and Path Analysis converged with the influence of causal analysis, all of which pervaded their thoughts on retention.

Spady (1975) used Durkheim's concept of integration as developed in Suicide to synthesize all previous research before the 1970's. Major points expressed in his research were: (1) bivariate research on the correlates of dropping out should be abandoned, and instead a multivariate statistical technique should be used, so "spuriousness" among the key variables could be identified as he argued that a path diagram would depict the dropout process much better: (2) that normative congruence, defined as a pre-condition for integration, between the students' attitudes, abilities, personal dispositions and the attributes and influences of the environment were essential to students' success in college: and (3) that the concept of satisfaction was a key variable

aside from social integration in influencing dropout decisions, as mediated through institutional commitment.

College student retention probably had become the area of research that had done the most to integrate the various administrative factors and academic disciplinary concerns in the research agenda of higher education. On the one hand, research in student retention was represented by the humanist approach that was exemplified in Attinasi's study of Mexican American student retention (1989). Symbolic interactionism and ethno-methodology, which have dominated the field of humanistic sociology, were employed in the exploration of the Mexican students' persistence on college campus. The underlying research perspective emphasized the context in which the students view the university as a relevant milieu and act on the meaning imposed on them. On the other hand, Spady and Tinto represented a positivistic approach to student retention. Labeling Tinto's approach as positivistic has historical origins, such as Durkheim's use of statistical materials in his study of sociology with its positivistic outlook. The theoretical reference adopted by Spady and Tinto in their approach to studying retention was largely based on Durkheim's statistical study of suicide in the late 19th century Europe. The concomitant variation statistical techniques used in Suicide (Durkheim, 1960) were in many respects similar to today's Pearson's correlation. The influence of Tinto's theory upon retention had been phenomenal; indeed, it has almost reached paradigmatic status according to Braxton (2000). Tinto's theory has become more nuanced and further developed. In fact, Tinto himself was a major critic of his original theory (1986) and has made the exposition of his theory even more clear. However, his criticisms as well as others only tend to concentrate on the verbal formulations of the theory without mentioning methodological difficulties and statistical issues. Thus, some of the conceptual difficulties and statistical issues in his theory have not been brought to the forefront. The main purpose of this paper is to explore Tinto's model in the context of causal modeling, its methodological difficulties, ramifications, and ensuing issues. Empirical studies using Tinto's theory were cited for illustration.

Tinto (1975) published his model, which in many ways was similar to Spady's. Under Tinto's theory, family background, personal disposition and schooling interacted with one another and this interaction ultimately had a direct influence upon goal commitment and institutional commitment. Goal commitment referred to the commitment to obtain a degree while institutional commitment referred to the individual commitment to a specific college. Goal commitment tended to have a direct influence upon academic performance while intellectual development had an influence on academic integration. Peer and faculty interaction reinforced each other and ultimately led to social integration (Tinto, 1975). Both social and academic interaction had an influence upon institutional and goal commitment which resulted in retention. In 1987, Tinto revised his model by inserting another variable, defined as 'intention to withdraw'. Figure 1 is the modified version. In his 1975 paper, Tinto specifically mentioned that the lines in his diagram did not necessarily represent paths in an interval path diagram. As a student of sociology, he was aware that the lines in his diagram could not be "paths" in an ordinary path analysis because the dependent variable of dropout was a dichotomous variable.

INSERT (Figure 1 here)

Tinto's difficulty stems from the fact that his diagram does not necessarily correspond to his verbal theory. In his diagram, Tinto posits that the relationship between academic integration and dropout behavior must be mediated by institutional and goal commitment. Thus, the relationship between dropout behavior and academic integration will vanish when the variables of goal commitment, institutional commitment, and intention are controlled. By this reasoning, the relationship between social integration and dropout behavior also requires institutional and goal commitment. Controlling the variable of institutional and goal commitment would erase the relationship between social integration and dropout behavior. According to the diagram, academic performance, faculty/staff interaction, extra-curricular activities and peer-group interactions have no bearing on the dropout decision when integration, intention and commitments are controlled.

However, one does not infer this conclusion from his verbal exposition. Rather, his verbal exposition takes into account numerous other factors. His verbal theory recognizes the relationship between constituent components of social and academic integration and dropout behavior that are unrelated to institutional and goal commitment. For example, he stated that experiences in the formal and informal system may also lead to voluntary withdrawal (Tinto, 1986). The confusion arises from the fact that in the diagram there is no direct path linking academic integration to dropout behavior, nor a direct path linking social integration to dropout behavior. While the discrepancies between Tinto's diagram and his verbal model constitute one of the larger problems of Tinto's theory, there are other issues that must be addressed. Other major issues are involved especially when Path Analysis is used with Tinto's model.

The first problem involves the lack of Operational Definitions. Tinto did not intend for his theory to be incorporated into Path Analysis; thus he did not provide operational definitions for his variables. Such operational definitions of the variables have been developed by Pascarella and Terenizi (1979, 1980) and further tested and refined by Cabrera, Castaneda, Nora and Hengstler (1992, 1996). The following operational definitions of each variable were excerpted from Pascarella and Terenizi (1986). Academic integration included (1) grade point average (GPA) for the freshman year; (2) satisfaction with intellectual development; (3) the student's perception of having a positive experience of intellectual growth. Social integration included (1) involvement in extra-curricular activities; (2) contacts with faculty; (3) having close personal relationships with other students; and, (4) interactions with faculty which had an influence on career goals and personal growth. Institutional commitment included: (1) confidence that the student made the right decision in choosing to attend this university. Goal commitment included the variables: (1) the highest expected academic degree; and, (2) the importance of graduating from college; (3) the rank of the enrolled institutions as a college choice; and, (4) confidence that the choice is the right one. Intent to persist included (1) the likelihood that the student would enroll at this university the following fall. All of these operational variables have the highest loading in the theoretical constructs. Munro (1981) defined satisfaction with faculty and satisfaction with work

skills as institutional commitment instead of academic integration in his study and found that it had no significant impact upon the dropout decision. Stage (1989) found that institutional commitment was related to withdrawal when satisfaction with social and academic life of the institution was defined as a major component of the variable of institutional commitment. Nora and Cabrera (1993) were more sophisticated in defining the variable of institutional commitment through their confirmatory factor analysis. Certainty of choice, institutional quality and institutional fit were included in their measurement of institutional commitment. The salient features of these operational definitions all point to the underlying concept of satisfaction. Most of the retention studies in the field of higher education, if not all, have adopted satisfaction as an approximation of either social or academic integration.

As Tinto's theory approaches near-paradigmatic status in the area of student retention, (Braxton, Milem and Sullivan, 2000), the contributions from Pascarella and Terezina, Cabrera, Castaneda, Nora, Hengstler, Braxton, Sullivan have been indispensable. Comparisons became possible only because there is a standard definition of each theoretical construct, which advances our understanding immensely. However, as all of these concepts are related to satisfaction, research will cast doubts on the validity of integration when satisfaction is found to be a suppressor variable in explaining college student retention (Bean, 1980).

The second problem involves the Part-Whole Correlation quandary or the substantive explanatory power of the variables. This question involves the variables of (1) intention to enroll: and, (2) enrollment status. The variable of intent to enroll the next semester should coincide with the real enrollment status of the student since the variable of intent is only a mental reflection of the overt behavior. In statistical jargon, it means that the intent to enroll is only an artifactual variable because the intent to enroll is a necessary precondition to enrollment status. Stated differently, enrollment status is a manifestation of the intent to enroll. This type of artifactual correlation is sometimes called part-whole correlation, indicating that the same observations are duplicated, in part, in the two series of measures (Muller, Schuessler and Costner, 1977). Since one has to assume that withdrawal is an intentional behavior, one withdraws because one intends to. Should they differ, it is because he/she changes his/her mind. If it does not entail any explanation of the withdrawal behavior, the variable of intent to withdraw is redundant.

The third problem involves Indirect and Spurious Effects when Path Analysis is used with Tinto's Model. Biostatisticians have long been critiqued for their tendency to ignore the problem of unmeasured sources of heterogeneity in longitudinal regression analysis. They tend to ignore these sources because they are primarily concerned about the effects of explanatory variables, and are not particularly concerned with testing hypotheses about the effect of time (Allison 1984). Psychologists, on the other hand, are noted for their concern towards casual ordering but not the decomposition of the effects. As Bentler (1980) noted, literature on the decomposition of effects aimed at attributing dependent variable variance to antecedent variables was generally ignored.

The well-known formula as cited in academic literature, such as in Duncan's paper (1966), rij= pij + Spik rjk where i and j represent two variables in the model and k

represents the whole set of the variables in the model with direct paths leading to the ith variable. For example r31=p31+ p32 r12 where i equal 3, j equals 1. Similarly r32 = p32 + p31r12, r42 = p42+ p43p32 + p41r12 + p43p31r12 and r43= p43+ p41r31+ p42r32. As seen in Figure 3) the equation r42=p42+p43p32+p42p12+p43p31r12, p42 is a direct effect, p43p32 is an indirect while both p41r12 and p43p31r12 are spurious effects. The relationship is spurious when the relationship between two variables is due to the antecedent variable, which in this example, is v1. In both terms, p42r12, and p43p31r12 involve the relationship with v1 and hence they are called spurious effects.

INSERT (Figure 2 here)

As in Psychology, researchers in higher education generally do not explicitly distinguish a variable's effects by its direct, indirect, spurious or non-causal nature. Most often only direct and indirect effects are reported, with the indirect effects including the spurious effects. When findings are presented in this way, only the direct effects are unambiguous. The indirect effects of the variables are exaggerated, while the effects of the antecedent variables are attenuated. Under Tinto's model (Figure 1), the effect of background skills is given less importance than it would be given if the spurious effect of intervening variables such as social integration and academic integration had been given appropriate recognition.

When the number of variables in the model becomes too large, the computation of indirect effects become tedious and is prone to error. Software to compute the indirect effects was first discussed and developed by Fox (1980, 1985). In his paper (1980), he explicitly defined the total effects as including only direct and indirect effects, and being exclusive of any non-causal or spurious effects. Although his approach is general enough to be applicable to most of the causal models, his software was written in APL, which has limited the wide use of this computer application.

The fourth problem involves the Causal Order of Variables, which should be considered when Path Analysis is used with Tinto's Model. Establishing a causal sequence out of regression coefficients in a model is generally considered highly speculative, yet in some cases such an exercise will remedy the theory's deficiencies. The example cited here is well known among practitioners in the fields of political science and sociology, and the principle involved is a repetition of the formula cited. When the postulate states that there is a path missing between the variables in the model, the regression coefficients of these variables computed from the data should be zero. In using this technique, Goldberg (1966) discerns the causal sequence of variables in voting behavior. His analysis, as shown in Model I, involves six variables: (1) the father's sociological characteristics (FSC); (2) the father's party identification (FPI); (3) the respondents' sociological characteristics (RSC); (4) the respondents' party identification (RPI); (5) the respondent's partisan attitudes (RPA); and, (6) the respondent's vote for president in 1956 (RV).

INSERT GOLDBERG'S (Model I) here

Prediction equations	Actual values
r41.23=0	.017
r61.2345=0	-.019
r32.1=0	.101
r52.134=0	.032
r62.1345=0	.053
r43.12=0	.130
r63.1245=0	-.022
r64.1235=0	.365

Having reviewed pertinent literature, Goldberg (1966) proposed a model in which some of the regression coefficients were zero (Model I). After the first revision of the original model, he proposed the voting behavior model as presented in Model II.

INSERT GOLDBERG'S (Model II) here

Prediction equations	Actual values
r41.23=0	-.017
r42.13=0	.357
r43.12=0	.031
r51.234=0	.037
r61.2345=0	-.019
r62.1345=0	.053
r63.1245=0	-.022
r64.1235=0	.470

INSERT GOLDBERG'S (Model III) here

Prediction Equations	Actual values
r41.23=0	-.017
r51.234=0	.083
r61.2345=0	-.019
r52.134=0	.032
r62.1345=0	.053
r53.124=0	-.073
r63.1245=0	-.022

Clearly substantive revisions are needed in Model II. A linkage between x4 and x6 should be inserted because the actual values of r46.1235 were .470. By the same

reasoning, a linkage between x2 and x4 should be inserted because the values of regression coefficient ($r_{24.13}$) were .357, which was well above zero as predicted.

Goldberg's approach had a wide influence on the application of causal modeling. Liu and Jung (1980) used the same technique to revise their satisfaction model to fit their data. In an exact identification model, Goldberg's technique is of no use because there is no predicted value of zero. Thus, an exact identification model is less interesting because data and model will have a perfect match and hence there is no room left for any improvement. Under Tinto's model, any predicted values of zero will lead to a revision of the original model in which some of the double arrows such as academic and social integration can be reordered in a sequential order.

The fifth problem involves Assessing fit between the data and the model when Path Analysis is used with Tinto's Model. The test for structural equation includes five steps: (1) model specification; (2) identification; (3) estimation; (4) testing fit; and, (5) re specification (Bollen and Long, 1992). Almost every researcher agrees that the last two steps are more controversial than the first three. In general, the chi-square test provides a test with the null hypothesis that the theoretical model fits the data. In a contingency table, one would like to see that the obtained chi-square is large and the probability value is small so one may conclude that the null hypothesis is false and the variables in the table are related. In Path Analysis, the converse is true. If the data fits the model, the chi-square should be small and the p-value should be large so that the null hypothesis is retained. However, the chi-square tends to grow larger when the sample size becomes large. In fact, the precise relationship between chi-square and sample sizes is given by Long (1986). When a study using Path Analysis involves a large sample, any trivial difference between the observed value and the model becomes significant and hence the null hypothesis is rejected. According to Mulaik et al. (1989), very good models were rejected because of the inadequacy of their chi-square test.

In order to supplement the chi-square tests, a variety of new measures have been proposed. Hoelter's CN is a modified chi-square test which has not received wide support. The main criticism of this measurement is that CN's variance tends to grow with the sample size (Bollen and Liang, 1988). Bentler and Bonett's (1980) normed-fit index (NFI) was another alternative to the chi-square test. Values of NFI can range from 0 to 1 and .9 is usually an indication of a good fit between the model and the data. Bollen (1989) proposed a refined version of this index. This new index adjusted the NFI for its sample size and degree of freedoms of the model. NFI applicability is even more useful when the problem of unknown statistical distributions in the NFI is alleviated through bootstrapping statistical techniques that can approximate the unknown statistical distributions (Bollen and Stine, 1992). The comparative index (CFI) given by Bentler (1990) provides another fit measurement regardless of sample size and its value is truncated to fall in the range of 0 to 1 where .9 indicates a very good fit. It is almost unanimous among social scientists that the fit of the model and data is controversial and that they should rely on various measurements instead of a simple chi-square test.

A fit between the data and the model is that R square is large and the NFI, CFI reflects the overall fit of the model to the data. A large square is probably easiest to

achieve when one just adds the variables into the model. The undesirable consequence of such an approach is that R-square is large enough to be significant, yet most individual regression coefficients are not. Thus, most researchers agree that the value of the t statistics for each regression coefficient should exceed 1.96, which is significant at .05.

The sixth problem involves consideration of the Suppressor variables when interpreting results. A renowned example cited about the issue of the suppressor variable was used in both Davis (1985) and Van de Geer's statistical texts (1971). In a numerical illustration of this relationship, data from Blau and Duncan (1967) is cited. They use four variables as follows: (1) variable X1=father's educational level: (2) variable X2=son's education level; (3) variable X3=occupational status of son's first job; and, (4) variable X4=occupational status of son's final job.

When the variable of x1, x2, x3 were used to regress on x4 the equation obtained the results as shown here: X4= -.X1+.8X2+.6X3++.734e (1.1). Since x1 is negatively related to X4, one would conclude that the father's educational level is negatively related to the son's later job. A naïve sociologist would conclude that in the 1970's, the hippie movement was a son's rebellion against his father's expectations. (Van de Geer, 1971). Thus, the son's later job appears to be independent of his father's education but depends on the son's education and his own first job status. Van de Geer, however, recognized that the positive effects of the father's educational level did not disappear but were merely absorbed by the additional variables of X2 and X3. He reasoned that in the same data it was found that the simple correlation between X1 and X4 is 0. In other words, the father's education level (X1) was a suppressor variable that was used to suppress the components in X2 and X3 in the equation (1.1), which were related to X4.

INSERT (Figure 3 here)

Similar findings in attrition were found in Bean's study. In his study, the variable of satisfaction was the cause of male student dropout (1980). His explanation was that men were satisfied with being in school, but were not studying hard. This plausible explanation was reasonable because he emphasized only the direct effect of satisfaction upon the decision to drop out. However, if he had adopted Van de Geers' interpretation, satisfaction would have been used as a suppressor variable that interacted with other variables in the model. In both men and women, a negative correlation was found between dropout and satisfaction, yet in regression coefficients, the relationship between dropout and satisfaction for men changed to positive. The variable of satisfaction, which contains the components of development, reutilization, and the Grade Point Average (GPA), was negatively related to dropout. However, the variable of satisfaction also contained components that were positively related to dropout. Thus, when the negative components were removed from the variable of satisfaction, only the positive components in the regression coefficient were left in the equation. A student might be very satisfied with the institution, which is quantitatively measurable, yet he might withdraw because of sudden changes in his family situation, which are embedded in the unmeasured part of the variable of satisfaction. A student may be well satisfied with every aspect of the institution; however, a housing shortage may preclude him/her from returning to campus. This residual part of the equation, which was not explained by the

negative components of routinization, grade point average and development, was a component, which might have a previously unexamined positive relationship with the dropout decision. Thus, the explanation that does not take into account the suppressor variable posits that the relationship between variable of satisfaction and dropout is negative. The shift in sign happened only when the relevant component was removed while other variables such as routinization, development and the GPA were controlled. In fact, with today's computer power, one can easily compute partial correlations between each pair of variables, such as satisfaction and routinization, satisfaction and GPA, with a control of a third or even fourth variable to reveal the intriguing relationship among the variables. Identifying the variable of satisfaction as a suppressor variable would lead a researcher to re-think whether integration measures should include a measurement of satisfaction. Its suppressing nature will attenuate its strength with the variable of dropout. This may be a plausible explanation for the non-significant relationship found between social integration and dropout, and academic integration and dropout in many retention studies in higher education.

Certainly, one might conclude that the different explanations by Van de Geer and Bean are a matter of style. In light of Bean's enormous contributions towards elucidating student dropout behavior, this is indeed a less significant issue. The explanation of suppressor variables is probably the least significant issue when discussing Path Analysis in this paper. When all the data such as simple correlation and partial correlation of the variables, are presented in the article, readers can make their own interpretations and are not necessarily swayed by the author's arguments. Van de Geer's interpretation of a suppressor variable is preferable to Bean's because Van de Geer's interpretation was based on the data at hand, while Bean's was based solely on his educated judgement.

The last problem involves the dichotomous nature of dependent variables when using Path Analysis or Multiple Regression procedures. The dichotomous nature of the dependent variable of retention/withdrawal has raised some statistical issues. Since Path Analysis is an extension of the ordinary regression analysis, all of the assumptions enacted in the regression analysis have to be observed in Path Analysis.

The dichotomous nature of the dependent variable of retention/withdrawal has raised some statistical issues. Since path analysis is an extension of the ordinary regression analysis, all of the assumptions enacted in the regression analysis have to be observed in path analysis. According to Hanushek and Jackson (1997), there were three major problems associated with the estimation of the regression coefficient. The first issue was that since there are only two outcomes of the dependent variable, the error terms could only assume two values. As a result, the error terms will vary with the independent variable, which makes the assumption of homoskedasticity untenable. The second issue is that the estimated value of the dependent variable will fall outside the range of [0, 1]. Any attempt to apply probability interpretation of the model becomes untenable when the predicted probability is beyond the [0,1] range. The third problem is a specification problem-- the model was assumed to be linear, yet it was non-linear.

The problem of homoskedasticity can be addressed by applying generalized least squares. Even assuming homoskedasticity, the most serious consequence of the use of

dichotomous variables is the yielding of a set of unbiased, although not the least variance, estimators. The second problem is even easier to resolve—it can be resolved simply by constraining the boundary between [0,1]. If one does not intend to use the probability function to predict some extreme values, this problem can be ignored. The third problem of specification is the most serious in that no feasible and easy answer can be found. If the nature of the model is non-linear, yet the function forms were assumed to be linear, the distribution of the observations becomes extremely important. If all the observations fall into the middle, no difference will be found among the linear and non-linear functions. Yet in the worst situation, the correlation in the non-linear model can be perfect but in the linear function the correlation is nil.

A few studies have explored the nature of this problem in real-life situations. Clearly and Angel (1984) have found that there were often insignificant differences between logistical regression, probit regression and ordinary regression. Dey and Astin (1991) drew the same conclusions when they used cross-validation techniques to replicate the original sample. However, Aldrich and Nelson (1985) have indicated that significant differences were found when different regression techniques were used in computing the same set of data.

Jorsekog and Sorbom in SPSS LISREL AND PRELIS (1989) indirectly discussed the issue of dichotomous variable. They suggested that the Likert scale was modeled as ordinal scale and that one could use PRELIS to estimate the parameters based on the assumption that the latent variable underlying the ordinal variable is continuous with the mean zero and unit variance. However, other studies have found that the problem of the dichotomous variable remains. For example, SPSS responded to a simulation study by Yung and Bentler (1994). The study found that a sample size of at least 2000, and possibly 5000, was needed to obtain satisfactory results (Smallwaters, 2001). By no longer providing polychoric/polyserial correlations in its statistical package, the problem of the dichotomous variable resurfaces.

The appropriateness of using the dichotomous variable in the equation depends on whether or not the function is linear. In general, a large sample with a moderate distribution (25% to 75%) of the dependent variable is approximately linear (Clearly and Angel, 1984, Goodman, 1975,) and the dichotomous dependent variable may not impede the path regression analysis. However, their estimations have not been subjected to rigorous testing. The guideline of a split of 75 % and 25% also presents empirical difficulties when institutions of higher education with moderately stringent admission criteria have a better than 75% of freshman to sophomore retention rate (Peterson's, 2001). Bootstrapping statistical techniques may provide further insight into the issue and hopefully will provide a definite solution to the problem.

Conclusions

In order to achieve an ideal model, one needs conceptually distinctive variables, which are generally statistically distinct as well. The approach adopted by Nora and Cabrera (1993) is especially meaningful because it will reduce the total variables of Tinto's model into fewer, yet conceptually distinct variables. Although there is no exact number of variables specified as desirable, the principle of parsimony is cardinal in Path Analysis.

Certainly all of these statistical issues have bearings on the validity of Tinto's model since every proposition stated in the model was empirically tested. In general, his model was well accepted in the field of higher education, yet quite a few of his propositions were not confirmed empirically. With so many variables embedded in the model, it becomes very difficult for all the propositions to be statistically significant. Simple mathematics explains this fact well: The more variables the model has, the lower the partial correlation each variable will have, since the total correlation cannot exceed one. With attention to the selection of the variables and the methodological issues raised in this paper, the model can be improved and more propositions will be acceptable.

If a metric increase in the independent variable is not a major concern, one can use many powerful statistical techniques to study retention. The logistical analysis advocated by Voorhees (1986), Liu (1980), and Liu and Sanders (1984) can be easily extended to the analysis of retention. Recent work by DesJardins and Pontiff (1999) indicated another approach to the study of retention. All of these approaches can reach the same conclusion as Path Analysis if one is not particularly interested in the differentiation of direct, indirect, and spurious effects of the variables.

Policy implications

Not all of the published findings of retention studies based on Tinto's model are equally important. The one variable that has received much attention is the variable of race. Race was often cited as not being a major variable in explaining the dropout behavior and appeared to have only minimal, if any, impact upon social or academic integration (Braxton, 1992). The findings of these studies may be attenuated because of the statistical issues discussed in this paper or because the samples chosen in the studies were too small. Many factors influenced the result of the study, none of which was more important than the appropriateness of the original theory. When the theory fails to provide adequate guidance for field research, the results of the study become obscure.

Tinto should not be responsible for the statistical problems associated with the model because he asked not to use Path Analysis to do retention analysis. Yet his theory of social integration and academic integration are open to critique. When Tinto borrowed the concept of integration from Durkheim, he also inadvertently accepted the assumption of normative congruence from Durkheim. For Durkheim, normative congruence has moral authority in that it leaves no choice for the individual. Regardless of whether one

was brought up in an Australian tribe or in the 19th century contemporary French society, normative congruence was imposed upon the individuals. This conservative nature of Durkhim's theory which reflected the educational philosophies of 19th century French Lycee was drastically different from the climate of today's American higher education.

Students with different cultural backgrounds were unlikely to accept normative academic congruence. Pascarella et al. (1996) found that white and non-white students differ in their attitudes regarding openness to diversity and multiracial challenges, which was expected. Surprisingly one of the reasons behind this difference was an institution's emphasis on being critical, evaluative and analytical. This emphasis had a positive effect on openness/diversity/challenge for white students, but a negative effect for their non-white counterparts (Pascarella, Edison, Nora, Hagedorn and Terenzini, 1996). Social congruence is also troublesome. Whether normative social congruence means integration to the whole student body or just to a sorority or fraternity is a question that needs to be answered.

Furthermore, Tinto's concepts of social and academic integration lack a rigorous frame of reference. Is integration a process or an outcome? Is integration a body of rules or regulations or a set of expectations? These questions will trigger a new set of questions, which may or may not be answered statistically. A symbolic interaction approach may be an alternative in studying student retention. The essence of this theory is that social organizations are not structured by univocal, and normative rules but by the reflections and dialogue by the actors themselves (Joas, 1987). Social relations are seen not as stabilized once and for all but as open and tied to ongoing common acknowledgement (Joas, 1987). Exchange and interaction is essential to the maintenance of the rules as well as alterations and reproductions of the rules.

Why does Tinto's integration deserve so much attention? His theory of integration is valuable because the issue of equal access of the seventies lingered into the nineties, and was further compounded with new issues of multiculturalism and diversity. Identity politics have become a central issue of campus politics. The Chicano student movement at UCLA, the African American students' movement at Rutgers, the American Indian students' protest at Michigan State (Rhoads, 1998) and Asian students' protest at Northwestern illustrated the fact that Generation X is no less demanding on the issue of equity than their parents were in the seventies. A profusion of literature in higher education indicates that the alienation of the ethnic minority student has been a major impediment to achieving their educational goals. (Loo and Rolison, 1986; Smedley, Meyers and Harrell, 1993). Statistics compiled by the National Education Statistics have noticed that the disparity in graduation rates among the European American students and ethnic minority students has not abated. All seemed to indicate that ethnic minority students have experienced problems in their adjustment to the college environment, which impede their academic and social success in college.

The relevance of Tinto's theory to today's campus culture is a paramount issue which researchers will likely have to address. The multicultural tenor of contemporary campus culture has great significance in the research of retention as well as policy formulation in general. When Stanford Law School flew all the admitted ethnic students

in for a campus visit, one can assume that this is not only for enrollment purposes. Along the same line, when Swarthmore flew in prospective ethnic minority students for a weekend stay, one has to realize that the idea of multicultural tenor has permeated into the policy formulation of elite institutions. It is on this basis that one has to question the relevance of Tinto's theory of integration to policy formulation. Tinto, a paramount figure, directed the field of higher education towards a paradigm of theory building. Along with him, much empirical research has led us to a point where we can have enough data to critically appraise the theory. Tinto's contribution may be limited to the studying of student retention, but he has changed the field of higher education from a field of study into a field of theory building.

Bibliography

Allison, P. D. *Event history analysis: Regression for longitudinal event data.* Sage University Papers: Quantitative Applications in the Social Sciences, Newbury Park, CA: Sage Publications, 1984

Attinasi, Jr L. C. " Getting In: Mexican Americans' perceptions of university attendance and the Implications for Freshman Year Persistence." *Journal of Higher Education*, 60 (1989), 247-277.

Bean, J. P., Dropouts and turnover: the synthesis and test of a causal model of student attrition, *Research in Higher Education,* 12(1980): 155-186.

Bean, J. P. Interaction effects based on class level in an exploratory model of college student dropout syndrome. *American Educational Research Journal*, 22(1985): 35-64.

Bentler. P. M., Multivariate analysis with latent variables: Causal modeling. *Ann Rev. Psychol.* 31(1980): 419-56.

Bentler. P.M., and Bonett., D. G. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, (1980): 588-606

Bentler. P.M., and Chou. C. P. Practical issues in structural modeling. *Sociological Methods & Research.*16 (1987): 78-117.

Bentler. P.M. Comparative fit indexes in structural models. *Psychological Bulletin*, 107(1990): 238-246

Bollen, K. A., A new incremental fit index for general structural equation models. *Sociological Methods & Research*, 17(1989): 303-316.

BEST COPY AVAILABLE

Bollen, K. A., and Liang, J., Some properties of Hoelter's CN. *Sociological Methods and Research,* 16(1988): 492-503.

Bollen, K A., and Scott, L., Tests for structural equation models. *Sociological Methods & Research*, 21(1992): 123-131.

Bollen, K. A., and Stine, R. A., Bootstrapping goodness-of fit measures in structural equation models. *Sociological Methods & Research*, 21(1992): 205-229.

Borgatta, E. F., and Bohrnstedt, G. W. A review of the background and history of the first fifteen years of sociological methods and research. *Sociological Methods & Research*, 17(1988): 3-20.

Braxton, J. M., Milem, J. F., and Sullivan, A. S. The influence of active learning on the college student departure process: Toward a revision of Tinto's theory. *Journal of Higher Education.*71 (2000): 569-615.

Cabrera, A. F., Nora, A., and Castaneda, M.B. College persistence: Structural equations modeling test of an integrated model of student integration. *Journal of Higher Education.* 64(1993): 123-139.

Cabrera, A. F., Nora, A., Castaneda., M.B. Nora., A., and Hengstler D. The convergence between two theories of college persistence. *Journal of Higher Education.* 63(1992): 143-164.

Clearly. P.D., and Angel, R. The analysis of relations involving dichotomous dependent variables. *Journal of Health and Social Behavior*, 25(1984):334-348

DesJardins, S. L., and Pontiff, H. Tracking institutional leavers: An application. *AIR Professional File*; n71 Spr 1999. 1999

Duncan, O.D. Path analyis: Sociological examples. *American Journal of* Sociology, 72(1966): 1-16.

Fox, J. Effect analysis in structural-equation models : extensions and simplified methods of computation. *Sociological Method & Research.*9 (1980): 3-28.

Fox, J. Effect analysis in structural-equation models II: Calculation of specific indirect effects. *Sociological Method & Research.*14 (1985): 81-95.

Golderberg, A. S., Discerning a causal pattern among data on voting behavior. *American Political Science Review*: 60(1966): 913-922.

Goodman, L. A. The relationship between modified and usual multiple-regression approaches to the analysis of dichotomous. Pp 83-110 in David Heise (ed.), Sociological methodology 1976. San Francisco: Jossey Bass.

Joas, H., Symbolic Interactionism in Gidddens, A and Turner J.H., edited Social *Theory Today* 1987 Stanford: Polity Press

Liu, R. A note on Phi-coefficient comparison" *Research in Higher Education*, 13(1980): 2-8.

Liu, R., and Loren. J. The commuter student and student satisfaction. *Research in Higher Education*; 12 (1980):215-26.

Liu, R., and Sanders, J. Selection of Models in 2x2x2 Contingency Tables. *Research in Higher Education*; 20(1984):17-21.

Long, J. S. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks: Sage Publications 1997.

Mooney, C. Z., and Duval, R. D., *Bootstrapping: A Nonparametric approach to Statistical Inference*, Thousand Oaks: Sage 1993.

Mulaik, S.A., James, L.R., Alstine, J. B., and Lind, S.D . Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105(1989): 430-445.

Munro, B. Dropouts from higher education: Path analysis of a national sample. *American Educational Research Journal* 20(1981):133-141

Friedrichs, R. W. *A Sociology of Sociology,* New York: Free Press 1972.

Muller, J. M., Schuelller, K. F., and Costner, H.L. *Statistical Reasoning in Sociology,* Boston: Hoghton Mifflin, 1977.

Namboodiri, N. K., Carter, L.F., and Blalock, H. M. *Applied Multivariate Analysis and Experimental Designs*, New York; McGraw-Hill,1975.

Nora A., and Cabrera F. The construct validity of institutional commitment: A confirmatory factor analysis. *Research in Higher Edcuaiton*:34 (1993):243-262

Peterson's. *4 Year Colleges* Peterson's: New Jersey, 2001

Pascarella, E. T., Terenzini, P. T., and Wolfle, L. M. Orientation to college and freshman year persistence withdrawal decisions. *Journal of HigherEducation*:57(1986):155 –175.

Pascarella, E. T., Edison, M., Nora, A., Hagedorn., L. S., and Terenzini., P. T. Influences on Students' Openness to Diversity and Challenge in the First Year of College. *Journal of Higher Education,* 67 (1996): 174-195.

Pascarella, E. T., and Terenzini, P. T. Predicting voluntary freshman persistence and voluntary dropout decisions from a theorecticl model. *The Journal of Higher Education,* 51(1980):60-75.

Rhoads. R. A. Student protest and multicultural reform: Making sense of campus unrest in the 1990s. Journal of Higher Edcuation.69(1998):621-644.

Smallwaters, P. Http://www.smallwaters.com/amos/faq/faqa-mfaq.html 2001.

Sobel., M. E. Direct and indirect effects in Linear Structural Equation models. Sociological Methods & Research:16(1987):155-176.

Stinchcombe, A.. *Constructing Social Theories.* Chicago: University of Chicago Press,1968.

Tinto, V. Dropout from Higher Education: A theoretical synthesis of recent research. *Review of Educational Research,* 45(1975): 89-125.

Tinto, V. *Leaving College: Rethinking the Causes and Cures of Student Attrition.* Chicago: University of Chicago Press, 1987.

Voorhees, R. A. Toward Building Models of Community College Persistence: A Log-Linear Analysis. AIR 1986 Annual Forum Paper. 1986

Wright, S. Correlation and causation. *Journal of Agric. Research* 20(1921):557-585.

XIE,Y. Structural equation models for ordinal variables: an analysis of occupational desintation. *Sociological Methods & Research,*17(1989):325-352.

Yung,Y. F., and Bentler, P.M. Bootstrap-corrected ADF test statistics in covariance structure analysis.47(1994):63-84.

Zilsel, E. Problems of Empiricism in *Foundations of the Unity of Science* edited by Otto, N, Carnap. R. and Charles M. Chicago: Universities of Chicago Press:(1968).

Figure 1

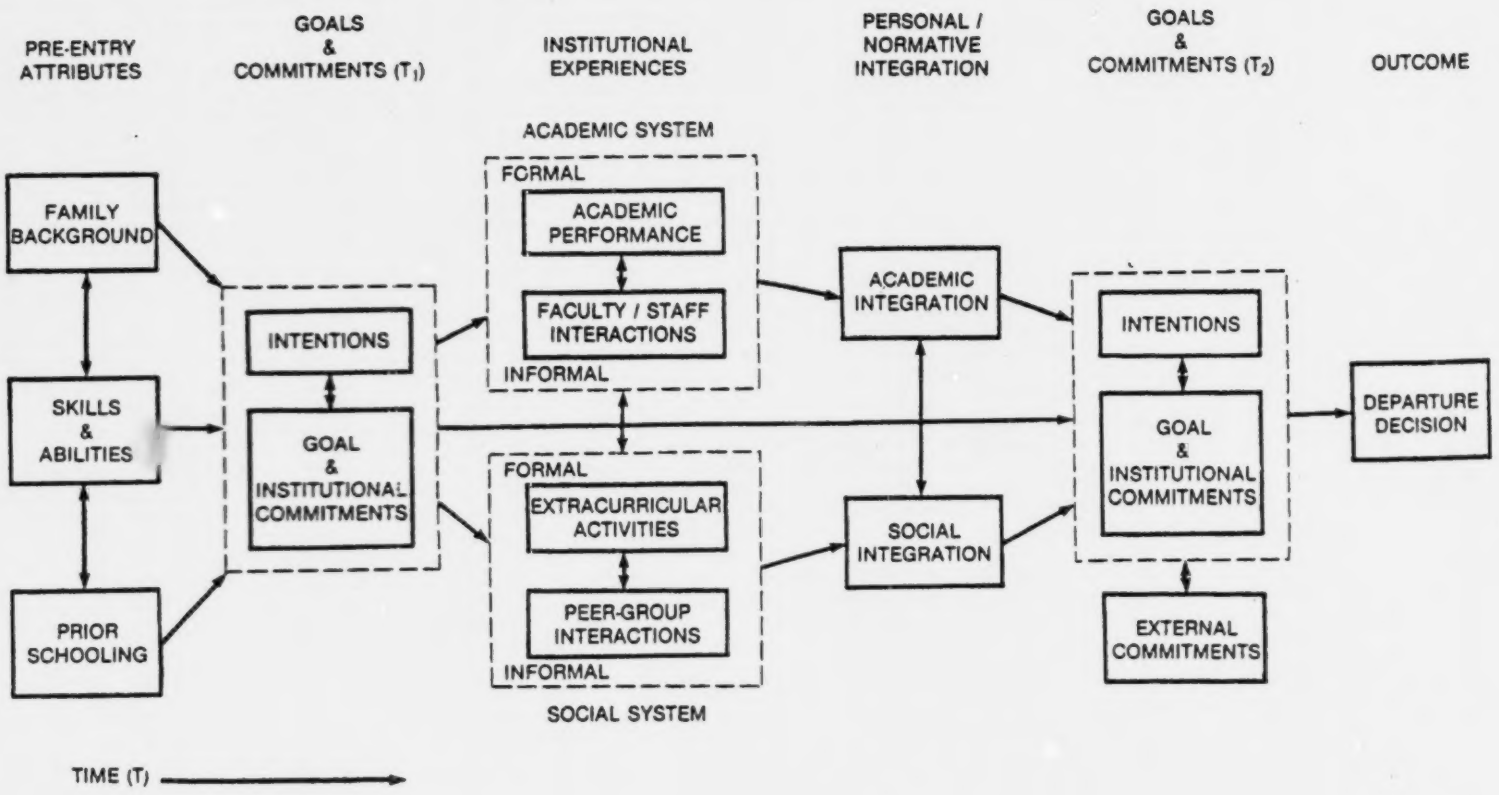
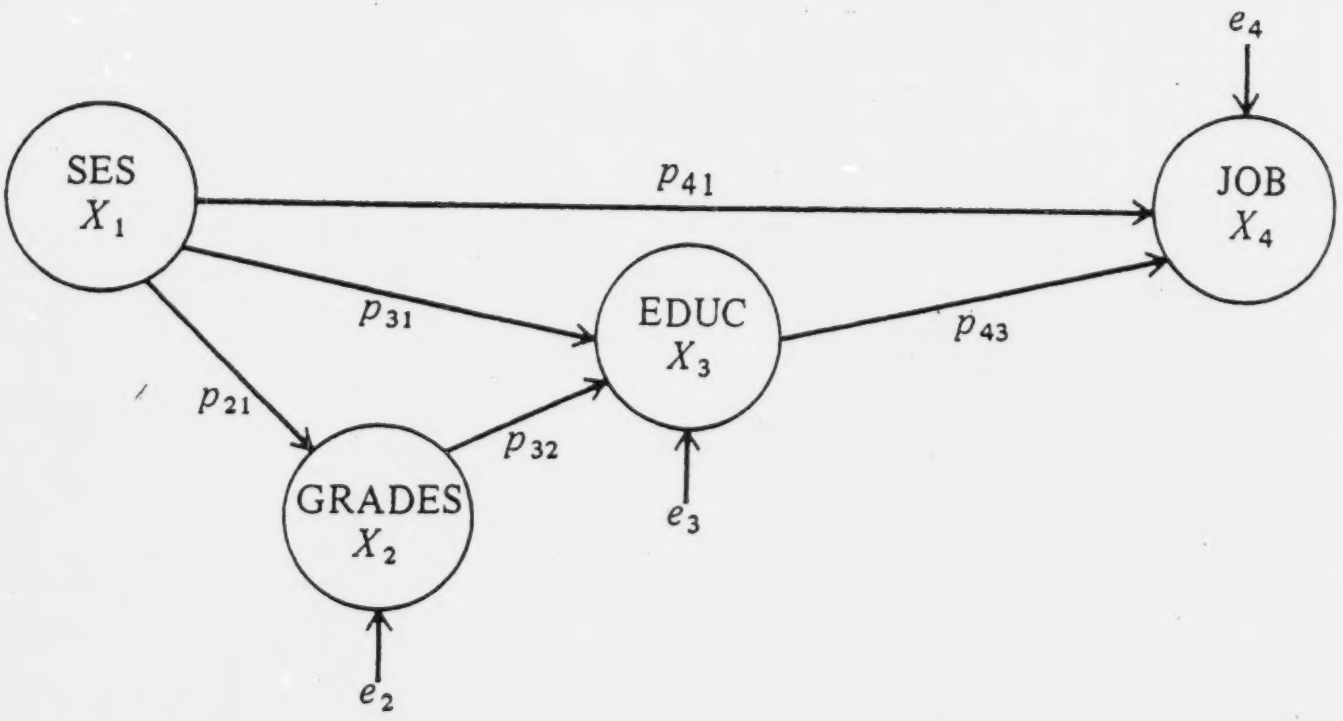


Fig. 4.1 A model of institutional departure

BEST COPY AVAILABLE

Figure 2



BEST COPY AVAILABLE

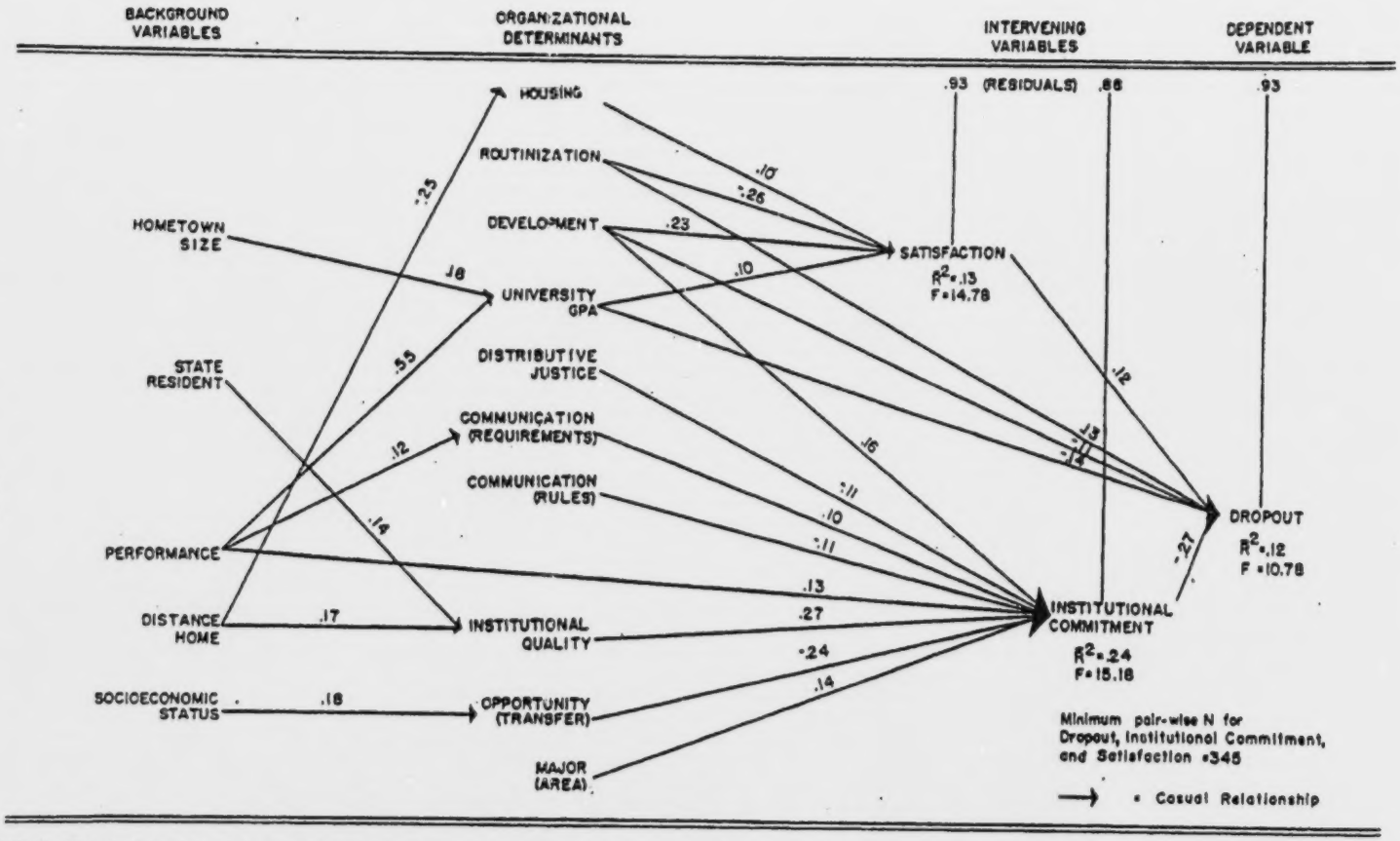
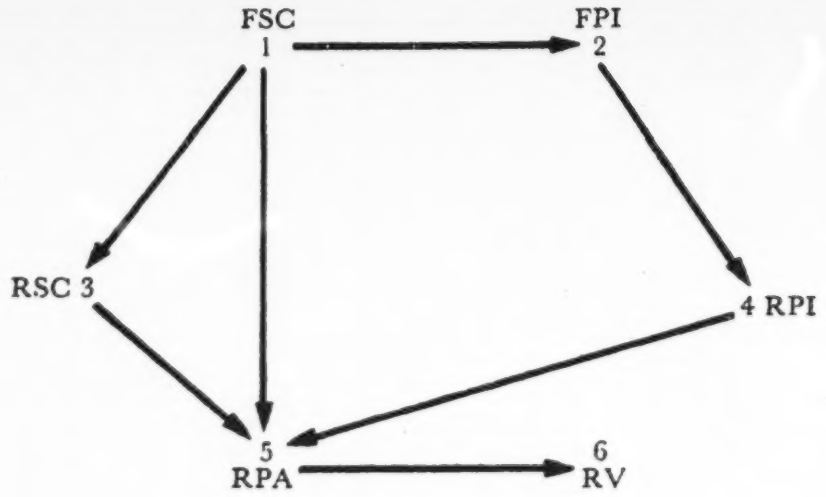


FIGURE 3. Path Model of Student Attrition for Men. Bean (1980)

Discerning a causal pattern among data on voting behavior

Model I. Attitude as final mediator.



Prediction equations	Actual values
$r_{41.23}=0$	0.017
$r_{61.2345}=0$	−0.019
$r_{32.1}=0$	0.101
$r_{52.134}=0$	0.032
$r_{62.1345}=0$	0.053
$r_{43.12}=0$	0.130
$r_{63.1245}=0$	−0.022
$r_{64.1235}=0$	0.365

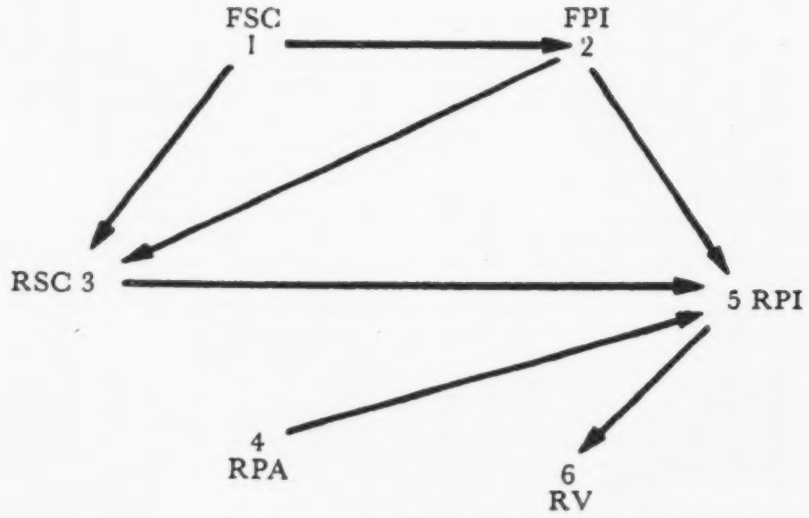
Figure 3.1.

BEST COPY AVAILABLE

Goldberg Model 2

Discerning a causal pattern among data on voting behavior

Model II. Party identification as final mediator.



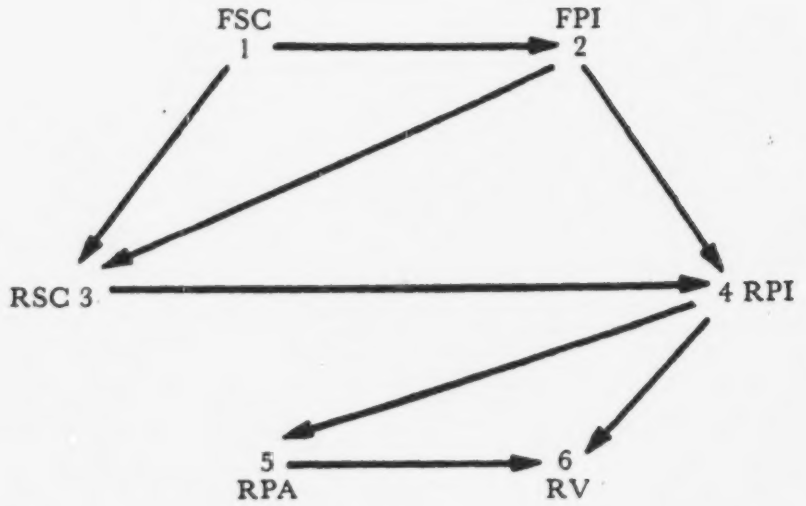
Prediction Equations	Actual Values
$r_{41.23}=0$	−0.017
$r_{51.234}=0$	0.037
$r_{61.2345}=0$	−0.019
$r_{42.13}=0$	0.357
$r_{62.1345}=0$	0.053
$r_{43.12}=0$	0.031
$r_{63.1245}=0$	−0.022
$r_{64.1235}=0$	0.470

Figure 3.2.

BEST COPY AVAILABLE

Discerning a causal pattern among data on voting behavior

Model III. Dual mediation.



Prediction Equations	Actual Values
$r_{41 \cdot 23} = 0$	−0.017
$r_{51 \cdot 234} = 0$	0.083
$r_{61 \cdot 2345} = 0$	−0.019
$r_{52 \cdot 134} = 0$	0.032
$r_{63 \cdot 1345} = 0$	0.053
$r_{53 \cdot 124} = 0$	−0.073
$r_{63 \cdot 1245} = 0$	−0.022

Figure 3.3.

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☑ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").